deep fake video

MARY MARKOVINOVIC

DKI APCSS

The makers

- Hollywood/gaming industry - in support of film visual affects (technology & content)
- Tech giants – enhancing technology for profit (technology/gaming)
- Consumers (Artists/fans) – for art or fun (content)
- Terrorists & Trolls– to spread fear (content)
- Government sponsored information operations (i.e., Russia and North Korea) – to undermine democracies and to intimidate (content & tech)

Fake imagery -How is it used?

Hollywood (and fans) have used CGI technology to:

- Put words into actor's mouths
- To insert characters into historical events

Bring back departed actors to finish roles

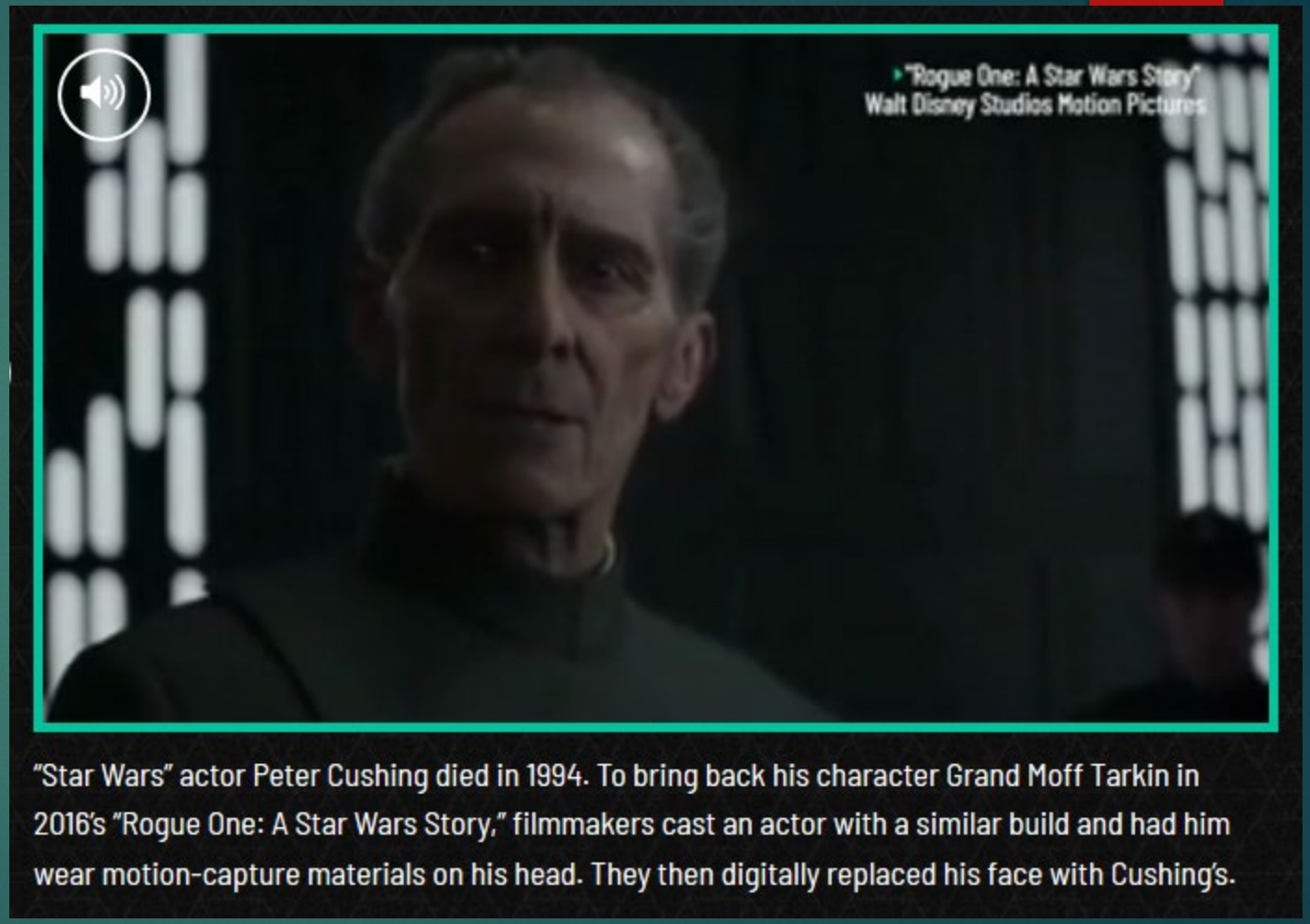
Adjust movements of characters/actors

For 1994's "Forrest Gump," filmmakers digitally inserted archival footage of JFK into this scene and manipulated his mouth movements.

How is it used?

Hollywood has used this technology to:

- To insert characters into historical events
- Put words into actor's mouths (language dubs)
- Bring back departed actors to finish roles
- Adjust movements of characters/actors



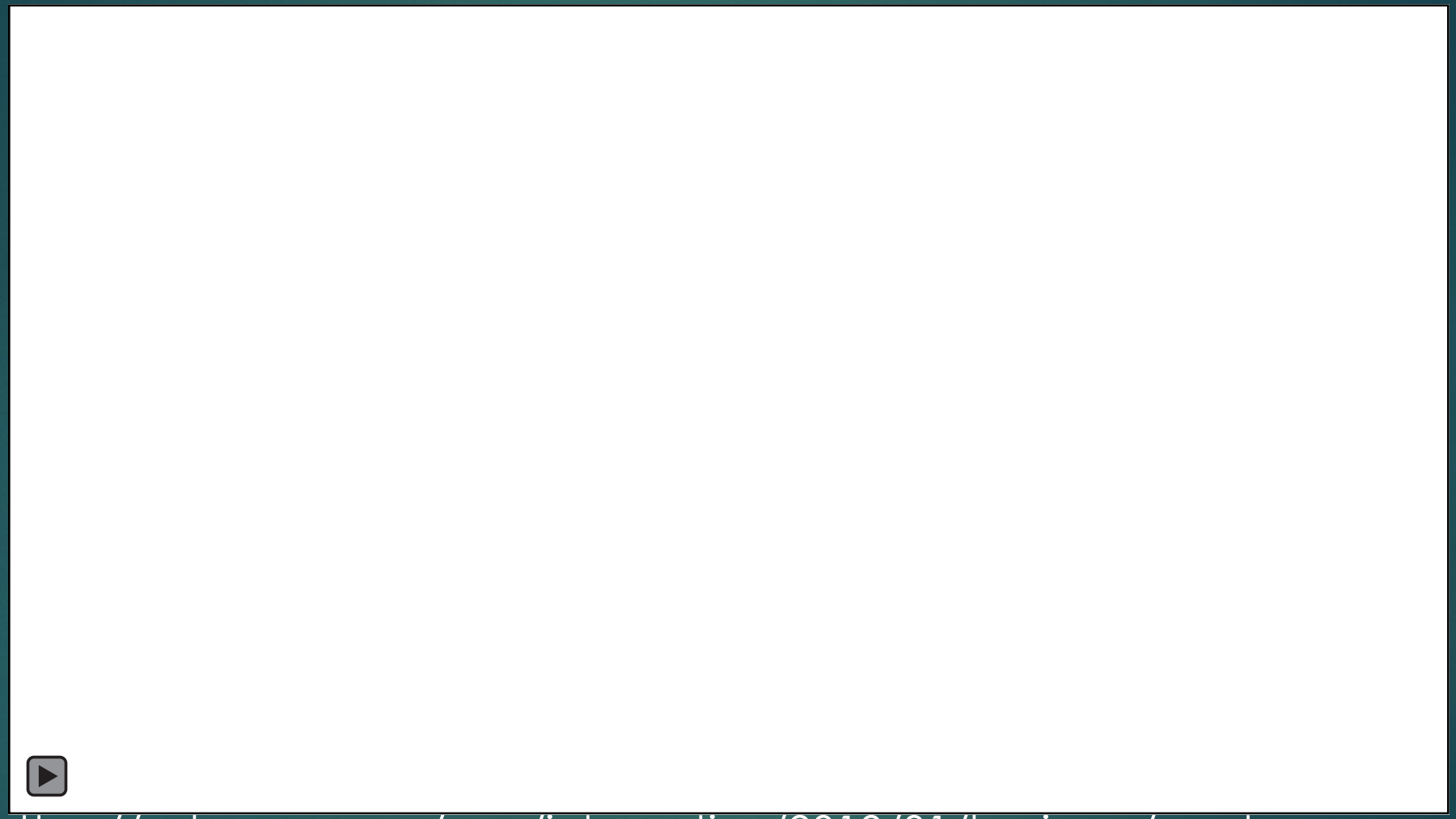
"Star Wars" actor Peter Cushing died in 1994. To bring back his character Grand Moff Tarkin in 2016's "Rogue One: A Star Wars Story," filmmakers cast an actor with a similar build and had him wear motion-capture materials on his head. They then digitally replaced his face with Cushing's.

CGI & Deep Fakes

- the capability to create altered audio and video files is no longer just for big budget Hollywood films. Its can be used by artists, content creators, businesses, and governments
- It's also used to create enhanced audio for things like "Siri" and "alexa"

"Robo-Journalism"

Example of deep fake video



- https://cdn.cnn.com/cnn/interactive/2019/01/business/pentagons-race-against-deepfakes/media/video/intro.mp4

What is this technology?

GENERATIVE ADVERSARIAL NETWORKS (GANS) USE ALGORITHMS TO BLEND PHOTOS, VIDEOS AND AUDIO.

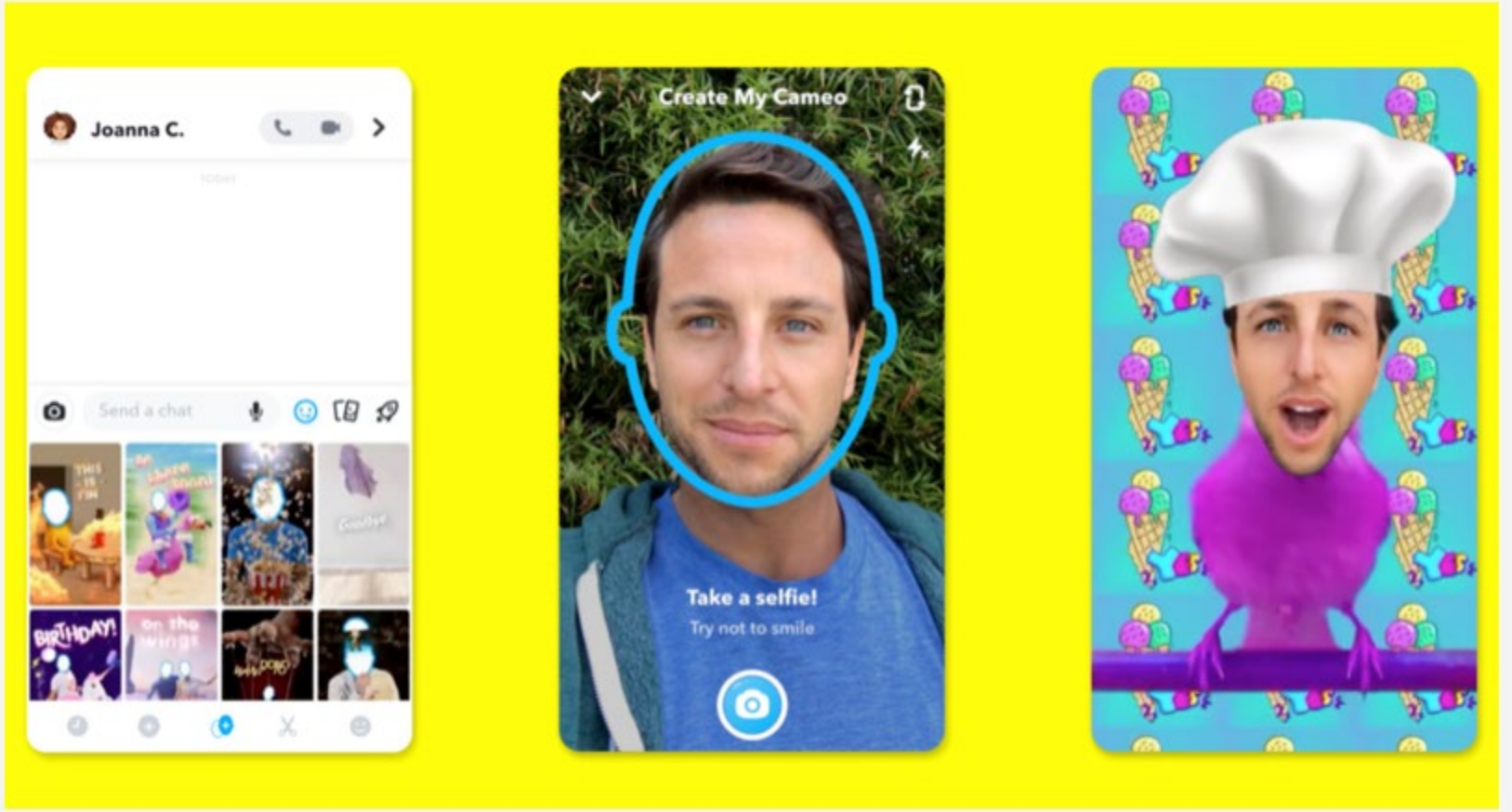
© Neurohive

Generative adversarial networks (**GANs**) are deep neural net architectures comprised of two nets, pitting one against the other (thus the "adversarial"). An algorithm is trained to recognize patterns in actual audio or visual recordings of a particular person, a process known as deep learning. As with doctored images, a piece of content can be altered by swapping in a new element — such as someone else's face or voice — and seamlessly joining the two. -

Can anyone access this tech?

- **Snapchat** filters – uses augmented reality to alter images. Easy to detect but continually being refined.
- **FakeAPP-** "an easy-to-use platform for making forged media.

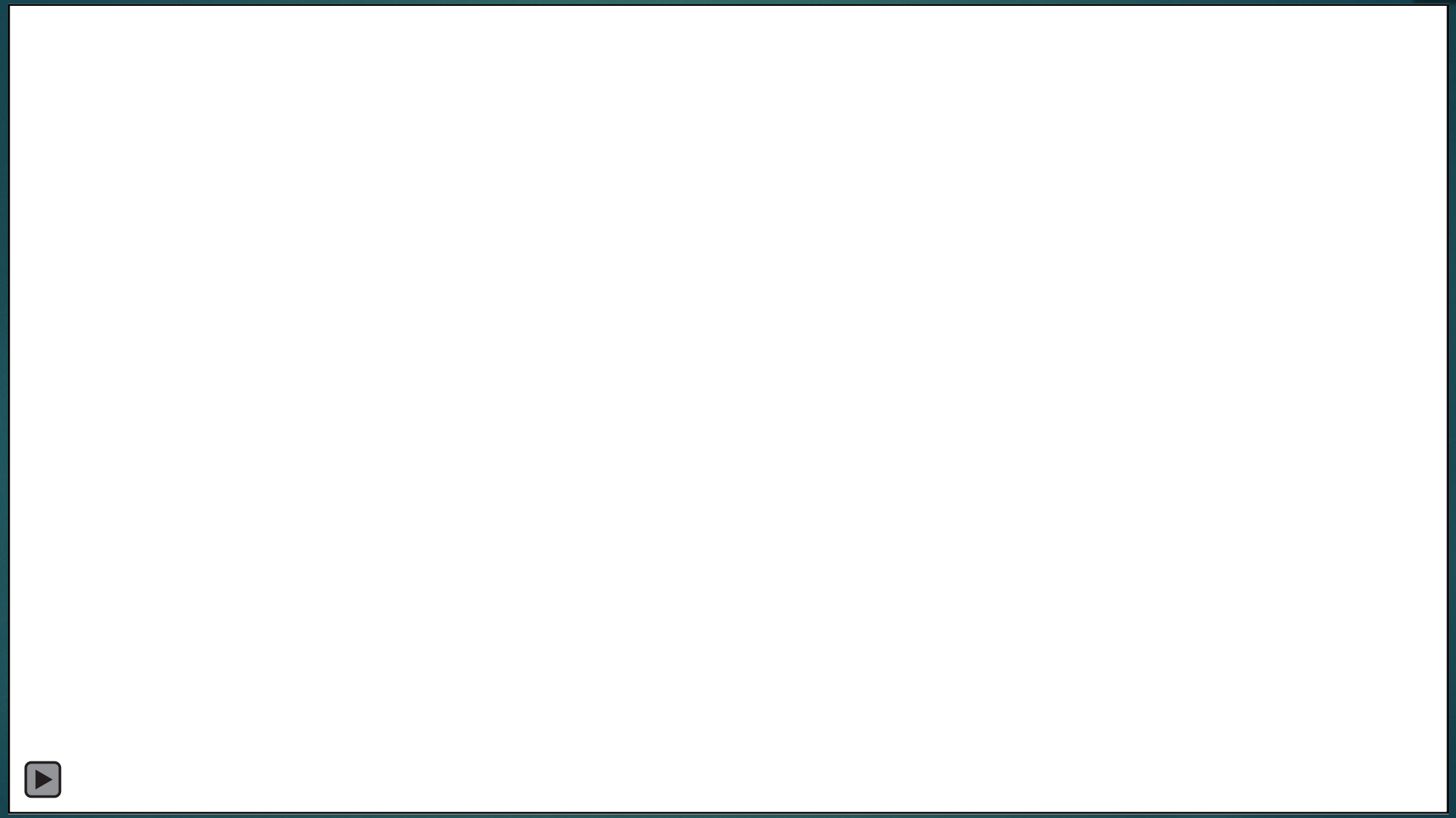
New Snapchat feature "Cameos" seamlessly edits users' faces into personalised videos and GIFs



#ZAO app

- Chinese Fake Video app - Released late August 2019.
- Enables users to Insert themselves in videos.
- Limited privacy controls.

Deepfakes example – Salvador Dali



https://youtu.be/64UN-cUmQMs

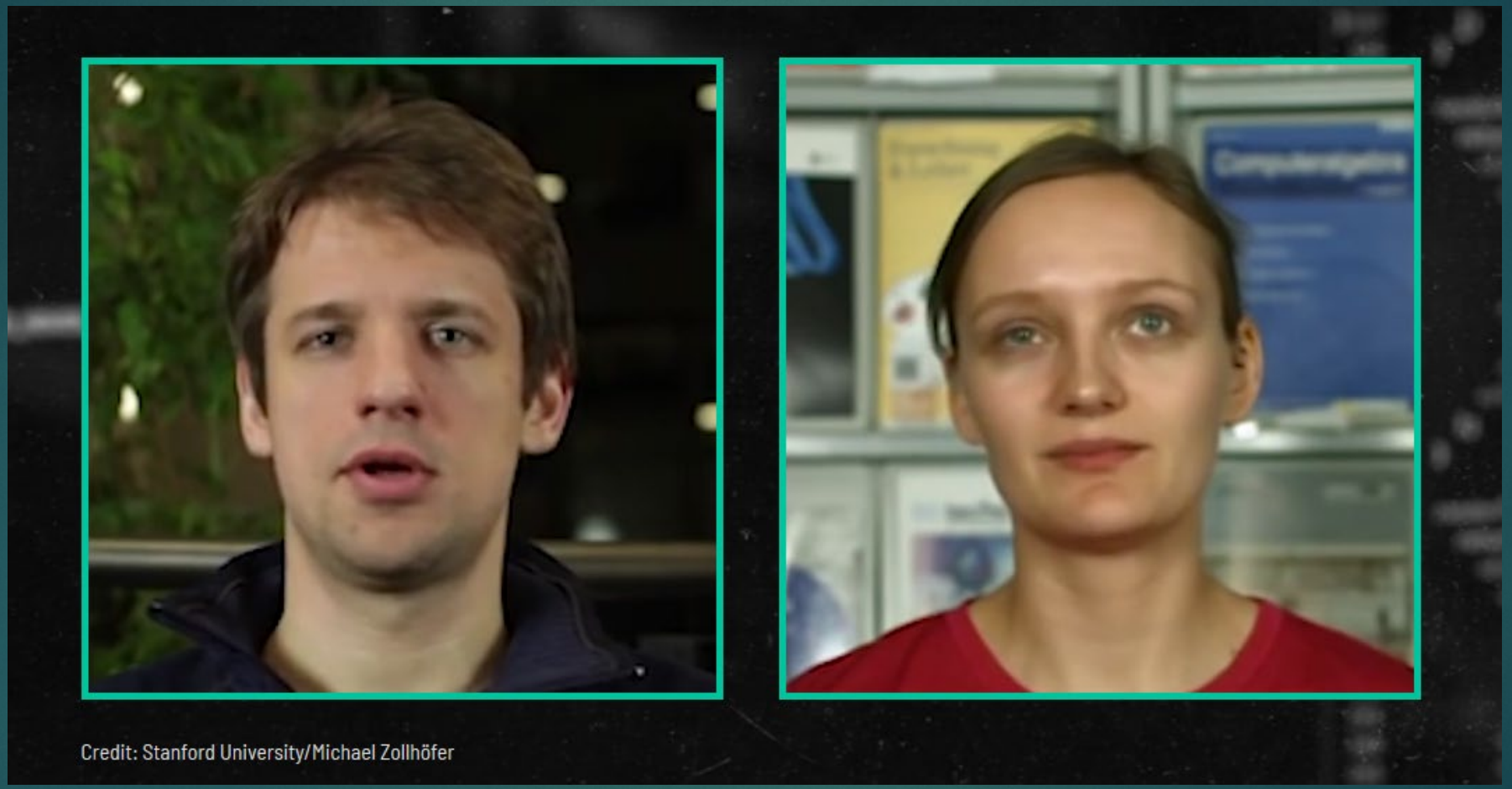
Why should we be concerned?

- As technology has become available and affordable to the general public there has been an increase in "swapping out heads" in videos.
- The number of DeepFake Videos online has exploded in the last year. Most victims are women.
- This technology could be "weaponized" for political or malicious purposes
- New business: creating DeepFakes or stopping them.

The Problem -- "Making a person appear to say or do something they did not has the potential to take the war of disinformation to a whole new level. " – Donie O'Sullivan, CNN

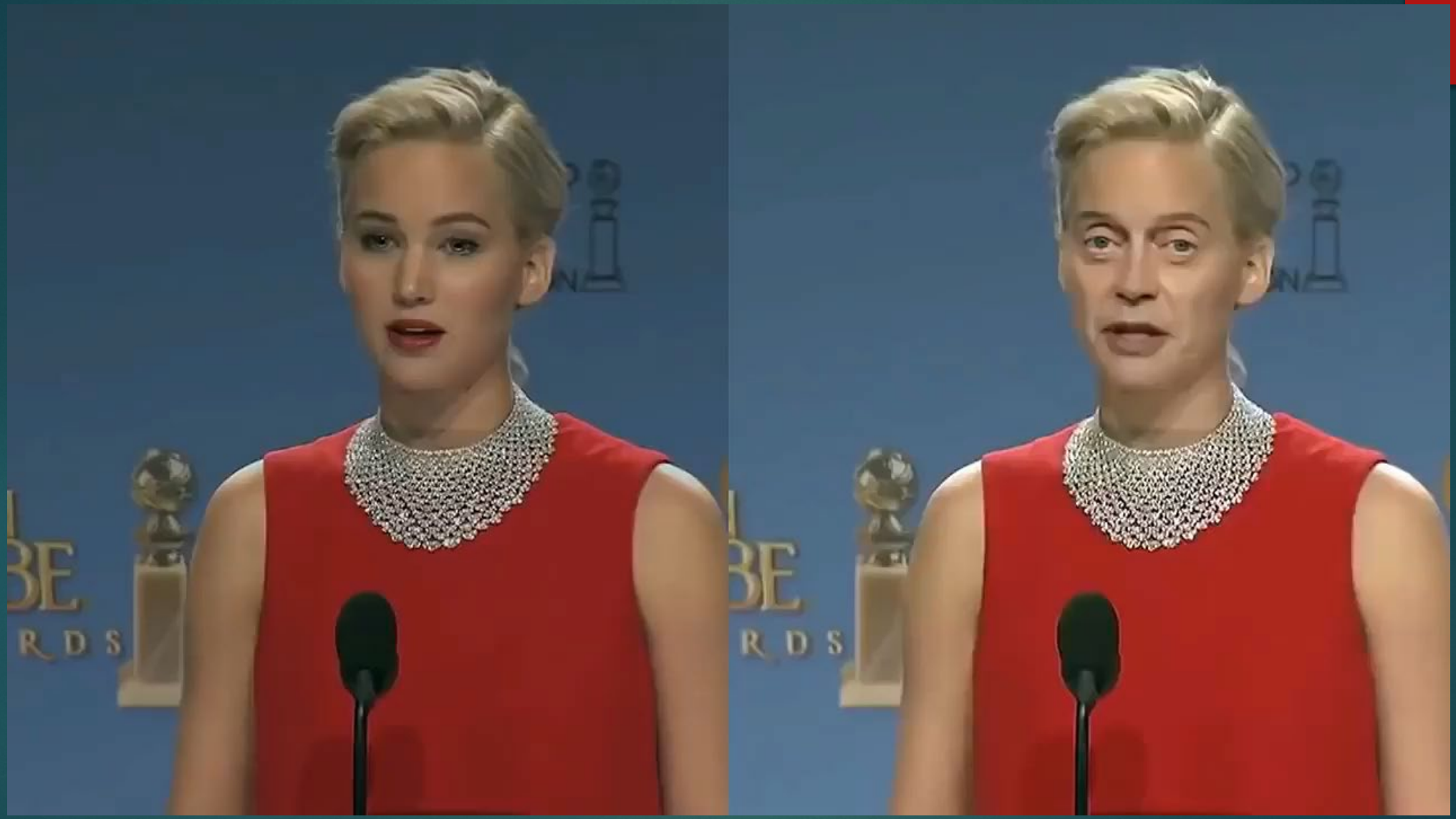
CNN----

Take the quiz: can you spot the deep fake



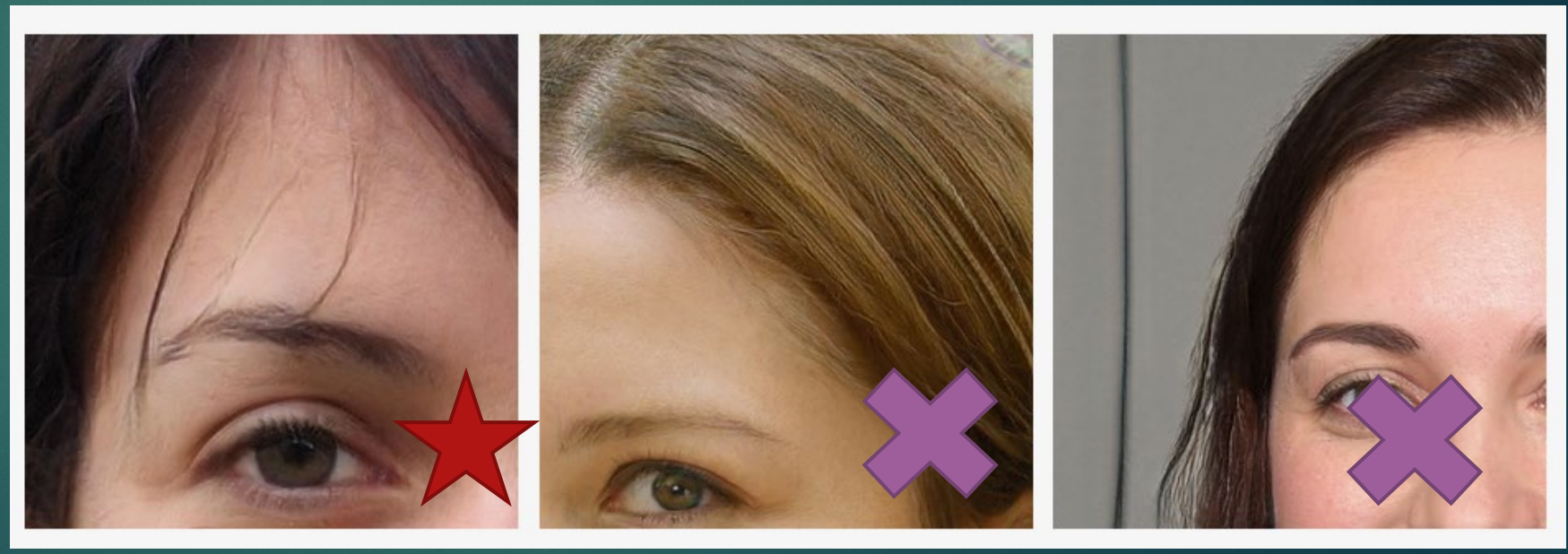
https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

From - CNN

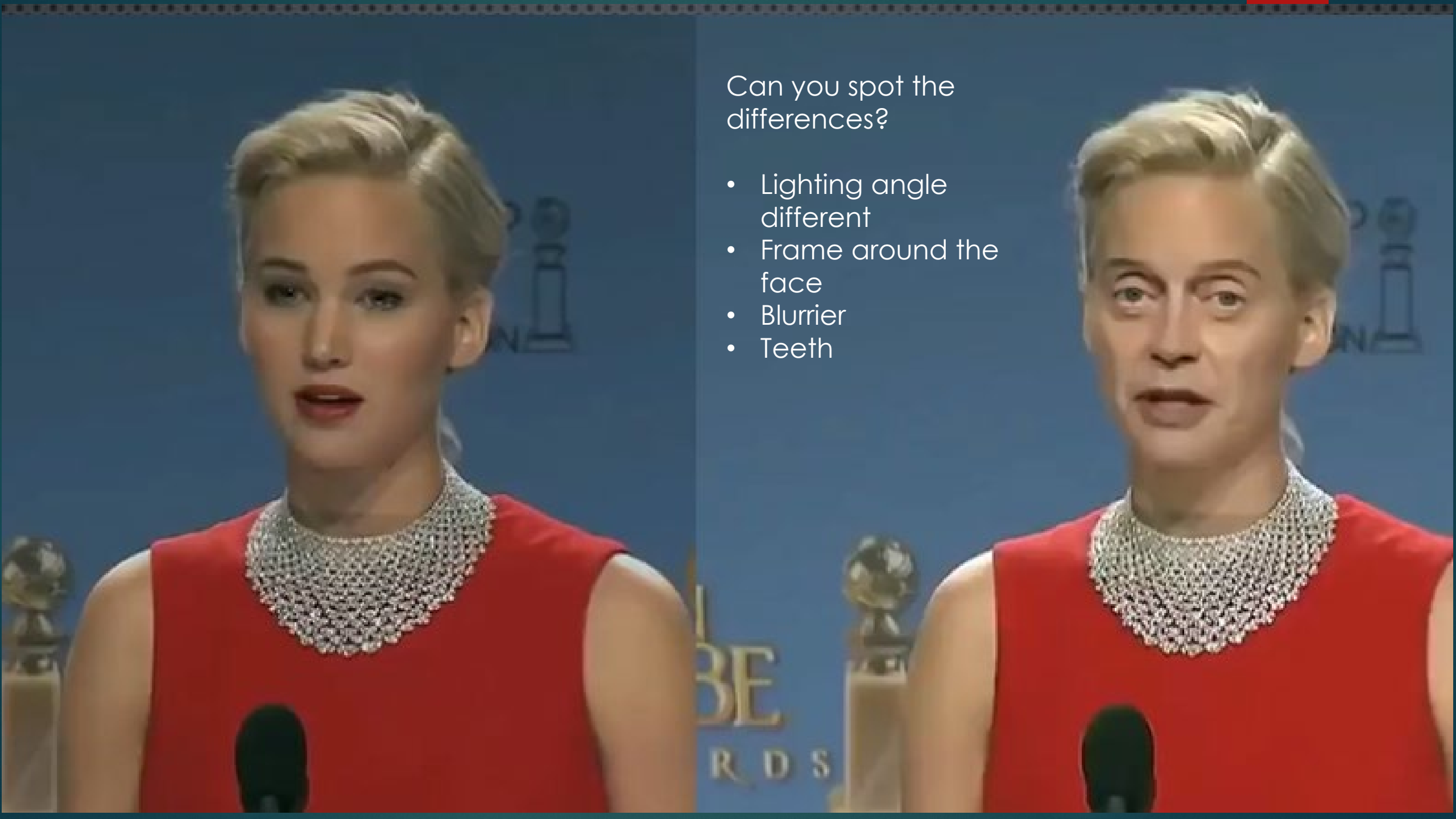


How is it detected?

- New technology is being developed to identify markers in videos which exist in fake videos.
- Here are some simple "tells" from http://www.whichfaceisreal.com/learn.html
  - Water splotches
  - Background problems
  - Eyeglasses
  - hair
  - lighting
  - Teeth
  - Other asymmetries



Hair is extremely difficult to render realistically



Can you spot the differences?

- Lighting angle different
- Frame around the face
- Blurrier
- Teeth

What happens if we can no longer trust what we see or hear?

- We already have this problem.

"The problem isn't just that deep fake technology is getting better. It is that the social processes by which we collectively come to know things and hold them to be true or untrue are under threat."

- Hany Farid UC Berkeley

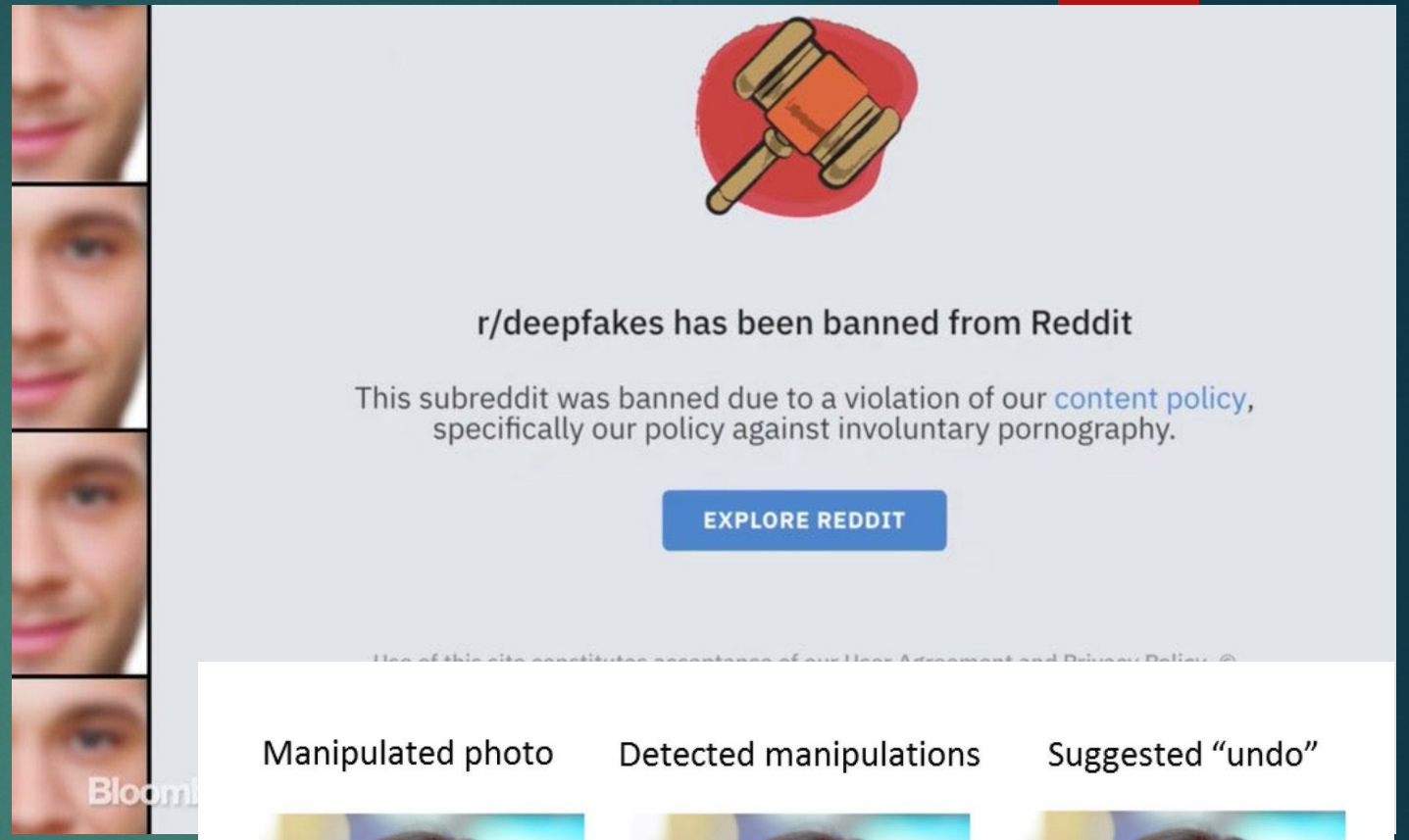
From - CNN

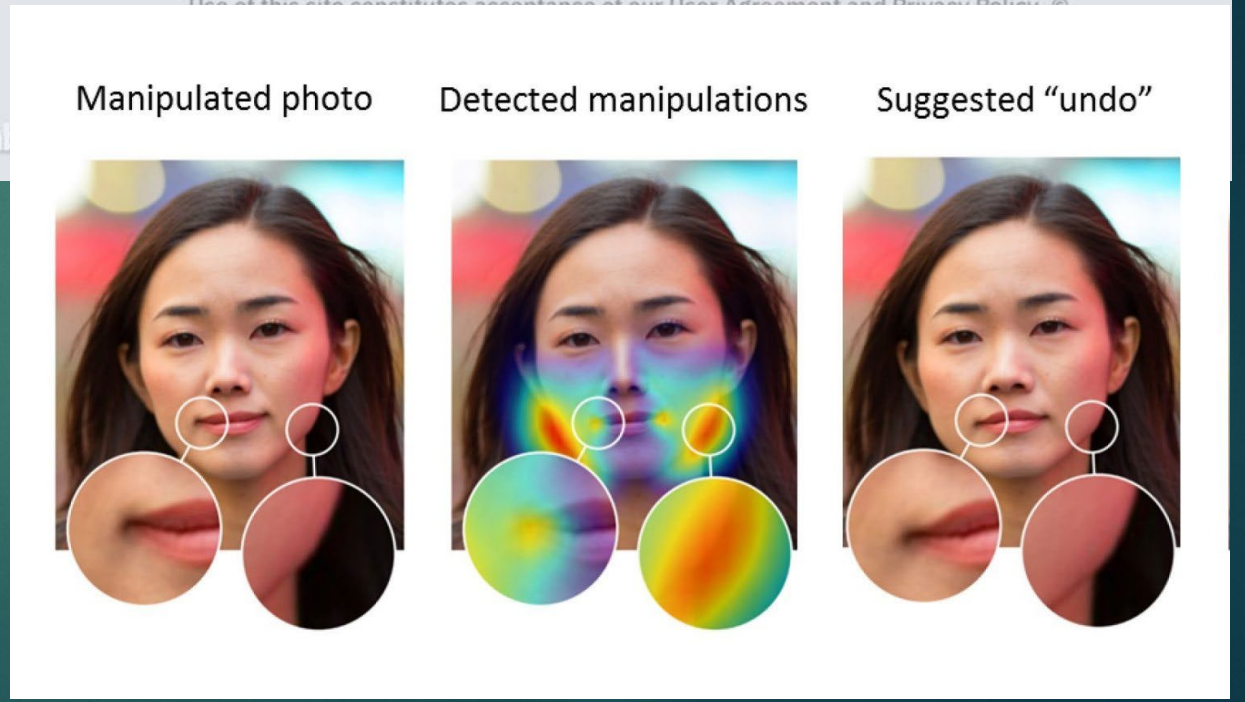
The emergence of deepfake technology has prompted members of the U.S. Congress to request a formal report from the Director of National Intelligence. Senator Marco Rubio worries about the global fallout after a convincing deepfake goes viral before it's detected.

From - CNN

Fighting deep fakes!

- Platform Policies
- Human reporting & intervention
- Algorithms & Artificial Intelligence





How can we stop deep fake videos?

DARPA is working with several universities to create detection software.

- Algorithms to detect
  - by coding
  - By physical motions (vowels/heartbeats)
- Algorithms to alter.

Defending Against Deepfakes

Discover SCIENCE FOR THE CURIOUS

The science of detecting deepfakes is, effectively, an arms race – fakers will get better at making their fictions, and so our research always has to try to keep up, and even get a bit ahead.

If there were a way to influence the algorithms that create deepfakes to be worse at their task, it would make our method better at detecting the fakes. My group has recently found a way to do just that.

At left, a face is easily detected in an image before our processing. In the middle, we've added perturbations that cause an algorithm to detect other faces, but not the real one. At right are the changes we added to the image, enhanced 30 times to be visible. (Credit: Siwei Lyu, CC BY-ND)

How can we stop deep fake videos?

- Some platforms such as reddit have limited bans. Expect to see platforms such as Facebook and YouTube start to implement systems to id. But what will the removal policy be? Who decides?
- Experts are calling for "digital watermarks" to be integrated into phone cameras.

- Companies such as Facebook, Adobe and Microsoft are developing DeepFake Detection Challenge

How can we stop deep fake videos?

- Facebook and Reuters are developing an online course to teach journalists to detect deepfakes.
- Alphabet created a new tool called Assembler to detect and analyze images and deepfakes.
- Governments are calling for a "Fake News" code of conduct for, or regulation of, Facebook, Google and Apple. (Deep Fake accountability act)
- Defamation, slander & Libel laws around the world need to be updated to better meet changing technology.

But once a fake video is out there, can the opinions of the people who saw it be corrected?

Improvements Continue

- As detection methods/systems improve so does DeepFake technology.
- Instant foreign language conversion and synchronization now available.

Nvidia researchers release "StyleGAN2" update that fixes visual flaws found in synthetic images generated by the previous model

Nvidia researchers released an updated version of their StyleGAN image generation algorithm that improves the overall quality of generated images and fixes previous unintended visual artefacts.

Latest advancements

Latest advancements

What can consumers do?

- Should we panic??
- No… we should make sure everyone is educated so that trust is not completely eroded.
- Social media makes worse the problem of confirmation bias. People need to gain better critical thinking skills or "intellectual resilience" to question propaganda or Fake News.

* http://www.independent.co.uk/news/uk/home-news/isis-winning-online-war-propaganda-extremist-material-radicalisation-report-facebook-twitter-social-a7954246.html

What can consumers do?

- Educate at all ages to help identify potential threats through media literacy
- Report malicious videos to platform moderators.

Questions?

By country

- The US House passed an annual national defence expenditure bill that includes the establishment of a $5m prize to stimulate R&D projects focused on deepfake detection technologies.
- California passed two state laws. AB730 prohibits the use of deceptive audio or video showing a candidate within 60 days of election. Must prove malice. AB602 prohibits the use of a persons image in sexually explicit material without their consent.
- Last week New York pass a state law protecting people from unauthorized used of their images in videos.
- China has a new Deep Fake Video Act
- Covered in Singapore Fake News law??